September 9, 2021

National Institute of Standards and Technology
100 Bureau Drive
Gaithersburg, MD 20899

BSA | The Software Alliance (BSA)[1] appreciates the opportunity to provide comments to the National Institute of Standards and Technology (NIST) on the draft document "Proposal for Identifying and Managing Bias in Artificial Intelligence" (Draft Proposal). BSA is an association of the world's leading enterprise software companies that provide businesses in every sector of the economy with tools to operate more competitively and innovate more responsibly. As companies at the forefront of AI innovation, BSA members are acutely aware of both the incredible potential that AI has to improve the world and its unique risks.

Bias has emerged as a top concern for policymakers, developers, deployers and the public as AI tools are used more and more frequently to make consequential decisions, like approval of loan applications or predicting healthcare outcomes. BSA recently released *Confronting Bias: A Framework to Build Trust in AI*,[2] a first-of-its-kind framework that organizations can use to perform impact assessments to identify and mitigate risks of bias that may emerge throughout an AI system's lifecycle. We hope that the BSA Framework will be a valuable resource as NIST works to create guidance around mitigating AI bias, as BSA sees significant overlap in the approach outlined in the BSA Framework and the Draft Proposal.

In particular, we would like to applaud several elements of the Draft Proposal that coincide with best practices laid out in our Framework. They include:

- **Emphasis on a lifecycle approach**. The BSA Framework identifies steps that can be taken in the design, development, and deployment stages of the AI lifecycle to mitigate the risk of bias. Bias can arise in a system at multiple points of its lifecycle and through many different channels, such as in the data used to train a model, in the formulation of the problem the system seeks to solve, or if a model is used in a scenario other than its intended purpose. Efforts to identify and mitigate the risk of bias in AI systems should therefore span throughout a system's lifecycle. We view a lifecycle approach as a fundamental starting point from which bias should be evaluated, and strongly agree with NIST's use of the approach in the Draft Proposal.

- **A comprehensive taxonomy of bias**. The inclusion of a comprehensive taxonomy of the types of bias that can emerge in an AI system further emphasizes the need

---

for a lifecycle approach to risk management. We outline many of these sources of bias in our Framework, and agree that the inclusion of a bias taxonomy will be useful for developers and deployers seeking to better understand the many ways bias can emerge in an AI system.

Overall, the Draft Proposal includes many of the key elements that BSA has identified as effective approaches for managing and addressing the risk of bias in AI, and we are in agreement with NIST on much of what is covered in the document. We offer below a few suggestions to further clarify aspects of the Draft Proposal to better convey the nuanced nature of managing the risk of bias, particularly in cases where it may implicate multiple stakeholders. NIST should consider:

- **Emphasizing that bias identification and mitigation may involve multiple stakeholders.** The Draft Proposal should account for scenarios in which the deployer and developer of a system are different entities and explain how responsibility for addressing bias may vary in those cases. Many BSA members provide B2B AI services to corporate customers that may retrain and/or customize the underlying AI model using their own data.[3] The Draft Proposal should account for such scenarios and explain how risk management responsibilities may vary depending on the nature of the AI system and the role that the deploying entity may have had in customizing the underlying model.

- **Clearer articulation of why bias must be treated as a risk management priority.** The underlying rationale of the Draft Proposal is that the risk of AI bias must be managed because it cannot be fully eliminated. We concur. Given the fundamental importance of this principle, we would recommend providing additional detail and context to explain why a risk management approach is necessary. For instance, the Draft Proposal could include a section to explain that effectively guarding against the harms that might arise from AI bias requires a risk management approach because: (1) the concept of bias – and corresponding definitions of "fairness" – are contextual and contested; (2) efforts to mitigate bias can involve tradeoffs that need to be evaluated on a case-by-case basis; and, (3) bias can emerge post-deployment, including in circumstances where a deploying entity uses the systems in an unforeseen or unintended manner.[4]

- **Recognizing the role of governance in AI risk management.** The Draft Proposal should highlight the critical role of that governance practices play in AI risk management. To that end, the Draft Proposal should emphasize that effective AI risk management should be underpinned by a governance framework that establishes the policies, processes, and personnel that will be used to identify, mitigate, and document risks throughout a system's lifecycle. In addition, a governance framework should promote understanding across organizational units—including product development, compliance, marketing, sales, and senior management—about each entity's role and responsibilities for promoting effective risk management during the design, development, and deployment of AI systems.

- **Evaluation of "fairness" metrics**. The concept fairness is an inherently contested ideal that is subject to multiple potential definitions. To manage the risk of AI bias it is nonetheless critical to select appropriate metrics for evaluating whether the

---

[3] See "Spectrum of AI Development and Deployment Models" on page 18 of Confronting Bias: BSA's Framework to Build Trust in AI, available at https://ai.bsa.org/confronting-bias-bsas-framework-to-build-trust-in-ai

[4] See "Managing the Risk of Bias" on page 9 of Confronting Bias: BSA's Framework to Build Trust in AI, available at https://ai.bsa.org/confronting-bias-bsas-framework-to-build-trust-in-ai

system is performing in an acceptable manner.[5] As Professor Arvind Narayanan famously noted, fairness can be mathematically represented using more than 21 different definitions that are impossible to satisfy simultaneously.[6] Given the critical importance that fairness metrics play in evaluating whether an AI system is performing in a manner that is unfairly biased, the Draft Proposal would benefit from the addition of a section that surveys the existing range of fairness metrics and discusses what factors that stakeholders can consider in determining whether particular metrics are relevant and/or appropriate for their use case.

- **Emphasizing documentation as a useful tool throughout the AI lifecycle**. Documentation can play a useful role in both the identification of AI bias risks and communication about how those risks have been mitigated. The Draft Proposal would be strengthened if it highlighted the types of documentation that can serve as useful artifacts for risk management activities, such as records of data provenance, documentation of stakeholder impacts, and risk mitigations.

- **Mapping the bias taxonomy to the relevant phases of the AI lifecycle.** NIST should add a section under each phase of the AI lifecycle that identifies the types of bias that may occur during that phase. Such a mapping would help stakeholders better understand how bias can emerge in a system and what tangible steps can be taken to mitigate those specific risks during each phase of the AI lifecycle.

BSA appreciates the opportunity to provide feedback on the Draft Proposal and looks forward to continued collaboration with NIST on this important topic.

---

[5] See page 19 of Confronting Bias: BSA's Framework to Build Trust in AI, available at https://ai.bsa.org/confronting-bias-bsas-framework-to-build-trust-in-ai

[6] Arvind Narayanan, 21 Fairness Definitions and Their Politics, ACM Conference on Fairness, Accountability and Transparency (March 1, 2018), https://www.youtube.com/ watch?v=jIXIuYdnyyk.